Proteins from Scratch

William F. DeGrado

Not long ago, it seemed inconceivable that proteins could be designed from scratch. Because each protein sequence has an astronomical number of potential conformations, it appeared that only an experimentalist with the evolutionary life span of Mother Nature could design a sequence capable of folding into a single, well-defined three-dimensional structure. But now, on page 82 of this issue, Dahiyat and Mayo (1) describe a new approach that makes de novo protein design as easy as running a computer program. Well almost. . .

The intellectual roots of this new work go back to the early 1980s when protein engineers first thought about designing proteins (2). At that point, the prediction of a protein's three-dimensional structure from its sequence alone seemed a difficult proposition. However, they opined that the inverse problem—designing an amino acid sequence capable of assuming a desired three-dimensional structure—would be a more tractable problem, because one could "over-engineer" the system to favor the desired folding pattern. Thus, the problem of de novo protein design reduced to two steps: selecting a desired tertiary structure and finding a sequence that would stabilize this fold. Dahiyat and Mayo have now mastered the second step with spectacular success. They have distilled the rules, insights, and paradigms gleaned from two decades of experiments (3) into a single computational algorithm that predicts an optimal sequence for a given fold. Further, when put to the test the algorithm actually predicted a sequence that folded into the desired three-dimensional structure. Thus, the rules of protein folding and computational methods for de novo design may now be sufficiently defined to allow the engineering of a variety of proteins.

Dahiyat and Mayo's program divides the interactions that stabilize protein structures

The author is in the Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6059, USA. E-mail: wdegrado@mail.med.upenn.edu

into three categories: interactions of side chains that are exposed to solvent, of side chains buried in the protein interior, and of parts of the protein that occupy more interfacial positions. Exposed residues contribute to stability, primarily through conformational preferences and weakly attractive, solvent-exposed polar interactions (4). The burial of hydrophobic residues in the well-packed in-

Better than the real thing. The natural zinc finger protein Zif268 (left) is stabilized in part by a core of hydrophobic (green) side chains and metal-chelating side chains (red). In the designed protein FSD-1 (right), the Zif268 core is retained but the metal-chelating His residues and one of the Cys residues of Zif268 are converted to hydrophobic Phe and Ala residues, thereby extending the hydrophobic core. The fourth metal ligand Cys^8 is converted to a Lys residue. The apolar portion of this interfacial residue shields the hydrophobic core, whereas its ammonium group is exposed to solvent. The helix is also stabilized by an N-capping interaction (19), which presumably also stabilizes the structure.

terior of a protein provides an even more powerful driving force for folding. The side chains in the interior of a protein adopt unique conformations, the prediction of which is a large combinatorial problem.

One important simplifying assumption arose from the early work of Janin et al. (5), who showed that each individual side chain can adopt a limited number of low-energy conformations (named rotamers), reducing the number of probable conformers available to a protein. This work was subsequently extended to the design of proteins containing only the most favorable rotamers (6). Although the side chains in natural proteins deviate from ideality in a few cases (complicating the prediction of the structures of natural proteins), these deviations need not be considered in the design of idealized proteins. Thus, various algorithms have been developed to examine all possible hydrophobic residues in all possible rotameric states, to find combinations that efficiently fill the interior of a protein. A complementary ap-

proach uses genetic methods to exhaustively search for sequences capable of filling a protein core (7), and this work has been adapted for the de novo design of proteins (8).

Interfacial residues are also quite important for protein stability (9, 10). They are often amphiphilic (for example, Lys, Arg, and Tyr) and their apolar atoms can cap the hydrophobic core, while their polar groups engage in electrostatic and hydrogen-bonded interactions.

Until recently, protein designers have frequently concentrated on quantifying the energetics associated with just one of these three types of interactions (3). However, de novo design is best approached by simultaneously considering all of the side chains in the protein—unfortunately, a very high-order combinatorial problem. For instance, the volume available to the interior side chains depends on the nature and conformation of the residues at the interfacial positions and vice versa. Dahiyat and Mayo assumed that each of these three features had been adequately quantitated to provide a useful empirical energy function for protein design. Their program combines a number of features taken from earlier potential functions and includes a penalty for exposing hydrophobic groups to solvent. Another essential innovation included in their program is an implementation of the Dead-End Elimination theorem, to efficiently search through sequence and side chain rotamer space.

Dahiyat and Mayo's target fold is a zinc finger, a motif with a well-established history in protein structure prediction and design. In an early, prescient paper, Berg correctly inferred that this His^2Cys^2 Zn-binding motif must feature a β-β-α fold that would position the ligating groups in a tetrahedral array around the bound Zn(II) (11). Favorable metal ion–ligand interactions together with a small apolar core help stabilize the three-dimensional structure of this compact fold. More recently, Imperiali and co-workers have designed a peptide that folded into this motif, even in the absence of metal ions (12). The design included a D-amino acid to stabilize a type II' turn, and a large, rigid tricyclic side chain that may help consolidate the hydrophobic core. This work was particularly ex-